7/PRTS

10/030652
531 Rec'd PCT/PTO 09 JAN 2002



70236/001.609

# SELF-ASSEMBLING BIOMOLECULAR STRUCTURES

This invention relates to self-assembling biomolecular structures, more particularly to such structures formed using modular units capable of biospecific interaction.

The assembly of ordered macromolecular protein structures is well known in nature, for example in the construction of transcriptional complexes in eucaryotes and in mechanically stabilising structures such as collagen in bone.

Peptides which contain alternating hydrophilic and hydrophobic amino acid residues and which are capable of spontaneous formation of macromolecular/macroscopic membranes are also known: see, for example, US-A-5670483. Such macroscopic materials, which may be held together by a combination of hydrophobic and ion-pair interactions, may be useful biomaterials for medical products such as slow release drug delivery vehicles, artificial sutures etc. It will be appreciated that the need to employ component molecules which are complementary and/or structurally compatible inevitably limits the application of such techniques.

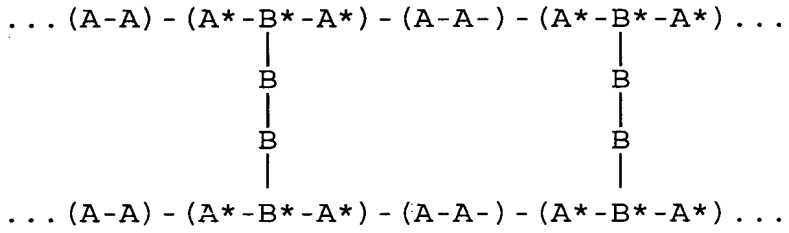
It has been described by Petka, W.A. *et al* in Science (1998) Vol. 281 pp 389-391 that hydrogels may be formed from a single class of multi-domain protein, here non-specific interactions are involved as unstructured aggregation occurs based on interhelical hydrophobic interactions.

The present invention is based on the finding that a wide variety of self-assembling biomolecular structures may be formed using self-organising modular units which are capable of biospecific interactions. Particularly of interest are protein structures formed using self-organising modular protein units. The

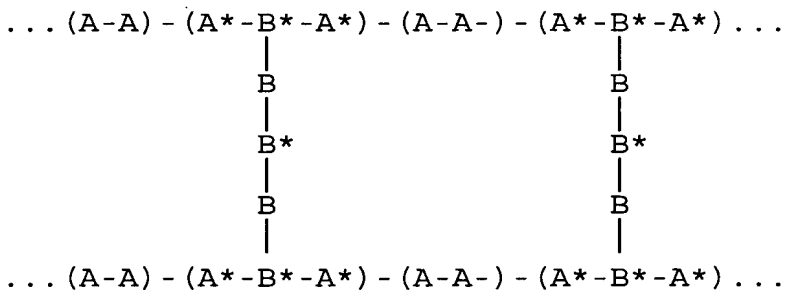
various modular units, hereinafter referred to as affinity modules, typically comprise appropriate combinations of protein molecules selected to convey the required capability for biospecific interaction. Thus, for example, where A and A* represent a first pair of proteins having specific affinity for each other (a first affinity pair), and B and B* represent a second pair of proteins having specific affinity for each other (a second affinity pair), affinity modules A-A and A*-B*-A* may be used to create self-assembling linear structures of the type

...(A-A)-(A*-B*-A*)-(A-A-)-(A*-B*-A*)...

Such linear structures only require a single affinity pair. B* is acting in this Example as a linker moiety for the two A* proteins. Additional use of B-B affinity modules may permit the assembly of sheet-like structures of the type




Three dimensional structures may similarly be assembled using an appropriate selection of affinity modules. Thus, for example, A-A, A*-B*-A* and B-B*-B modules may self-assemble to form sheets of the type



Two such sheets may be linked together by using B-B modules to bind the B* moieties in the B-B*-B linking groups of adjacent sheets, so as to form a two-sheet structure.

It will be appreciated from the foregoing that by using two or more affinity pairs of proteins (which term as used herein with regard to the invention also embraces appropriate peptides, polypeptides, protein domains and fragments etc.) which have specific and exclusive affinity for each other, by choosing appropriately constructed affinity modules in appropriate relative amounts, and where necessary by interacting the affinity modules in an appropriate order, it is possible to form self-assembled protein structures in a highly specific, predictable and controllable manner and with widely ranging degrees of complexity.

Thus according to one aspect of the invention there is provided a self-assembled protein structure comprising at least two pairs of proteins (affinity domains) wherein the proteins of each of said pairs have specific and exclusive affinity for each other, said proteins being linked to form affinity modules capable of mutual biospecific interaction to form an assembled structure.

More generally, and having applications which are discussed in more detail below, the self-assembled structure may comprise non-proteinaceous components and thus in a further aspect, the present invention provides a self-assembled biomolecular structure comprising affinity modules, said affinity modules each having at least two affinity domains which may be the same or different, at least one affinity domain within each affinity module having specific and exclusive affinity for an affinity domain within another affinity module, said affinity modules capable of biospecific interaction to form an assembled structure.

Preferably, the self-assembled biomolecular structure will comprise at least two pairs of affinity domains, each of said pairs having specific and exclusive affinity for each other. In one preferred embodiment, each affinity domain within a given affinity module will have specific and exclusive affinity for another affinity domain.

Preferably, the self-assembled structure will comprise at least two different types of affinity modules. Preferably at least one of the affinity modules will have two or more different affinity domains. In a particularly preferred embodiment, the structure comprises at least one affinity module which has two or more functionally equivalent affinity domains and at least one affinity module which has two or more different affinity domains.

Preferably the affinity modules and affinity domains will be proteinaceous but the self-assembled structure may incorporate non-proteinaceous molecules capable of specific interactions with other molecules (typically proteins) such as polynucleotides, polysaccharides, haptens, vitamins, lipids etc. Thus, one or more of the affinity domains or affinity modules may comprise such a non-proteinaceous molecule. Affinity domains will be considered 'proteinaceous' even though they may contain a non-protein part such as a co-factor. In one preferred embodiment, affinity modules which comprise affinity domains allowing mutual recognition will also comprise domains capable of binding to a non-proteinaceous ligand such as a polynucleotide, polysuccaride etc. as mentioned above. Affinity modules can be designed in such a way that the full self-assembling structure is only formed in the presence of this non-proteinaceous ligand, thus providing a useful diagnostic tool where the self-assembled structure can be readily detected.

Such a technique is also applicable for detecting

proteinaceous ligands, affinity modules being designed with domains which are capable of specific interaction with the proteinaceous ligand of interest which is then incorporated into the self-assembled structure. Alternatively a ligand, which need not be proteinaceous, could be encapsulated by interaction with affinity modules for therapeutic use, e.g. drug delivery, possibly in a controlled release system.

The affinity domains are identifiable molecular units having the binding affinities discussed herein. The term 'domain' is used functionally to refer to a unit having a certain binding affinity and does not imply a particular structural arrangement. The affinity domains may be naturally occurring molecules or fragments thereof or such molecules which have been modified or molecules designed de novo. Preferably, at least one of the affinity domains is an 'affibody', a protein domain which has been engineered to have a new or enhanced binding affinity for a target molecule. Engineering may be by combinatorial engineering or possibly by rational design.

The present invention relates only to non-naturally occurring self-assembled structures. While one or more of the affinity domains or affinity modules may be naturally occurring (although conveniently prepared by recombinant techniques) the overall structures are not the same as those which are formed in nature. Preferably, one or more of the affinity domains will have been selected from a molecular library. As described in more detail in the present Examples, a domain which has good affinity for a molecule which is to form another affinity domain within the self-assembled structure can be selected from a displayed molecular library. Molecular libraries can be prepared by combinatorial engineering of a given protein or protein domain and candidate affinity domains selected. Alternatively, libraries can also be constructed by

harvesting collections of genes from natural sources, which have already in a sense been randomised (Hoogenboom *et al.*, Immunotechnology **4**, pp. 1-20 [1998]). Convenient affinity domains include bacterial receptors, domains derivatives and fragments thereof.

Each affinity domain thus has its own specific binding partner, i.e. an affinity domain with which it enjoys exclusive and specific interaction, forming an affinity pair. In a given self-assembled biomolecular structure, affinity modules may comprise affinity domains which are not able to take part in the specific interactions because a suitable binding partner is not present. For example in Examples 4 and 5 ABD is not able to undergo specific interactions with an affinity partner and the structure depends only on interactions between the $Z_{8G}$ domains and SPA. A further type of affinity module could be provided which incorporated, for example, an albumin affinity domain and the structure would then comprise 2 affinity pairs, 2 pairs of specifically interacting affinity domains.

Reference is made herein to "specific and exclusive affinity" between affinity domains. The terms does not imply that a given affinity domain could only ever interact with a single molecular species, rather, within the system each affinity domain has a clearly identifiable affinity partner with which it interacts much more strongly than any other domain present. The affinity domains $Z_{8G}$ and $Z_{6S}$ described in the Examples both have 'specific and exclusive affinity' for SPA within the meaning of the present invention, i.e. each is capable of forming an affinity pair with a domain of SPA.

The interactions may be based on contributions from a number of different types of interactions including hydrogen bonds, these interactions typically being strongly sequence and conformation dependent. The affinity between domains is specific, rather than

relying on general interactions based on broad functional characteristics such as hydrophobicity which could be displayed by a large group of molecules. Typical of the specific and exclusive affinity observed between affinity domains in the present invention are protein:protein, antigen:antibody type interactions.

The aforesaid affinity modules, which themselves constitute a feature of the invention, may for example comprise naturally small and highly soluble protein affinity domains, such as individual domains of the bacterial surface receptors of staphylococcal protein A (SPA) or streptococcal protein G (SPG) (Ståhl, S. and Nygren, P-Å., Patologie Biologie **45**, pp. 66-76 [1997]; Jansson, B. *et al.*, FEMS Microbiology and Immunology **20**, pp. 69-78 [1998]). Domains having affinity therefor may be found in collections or libraries of individual domains of SPA (Nord, K. *et al.*, Nature Biotechnology **15**, pp. 772-777 [1997]), SPG, antibody fragments such as single-chain Fv, Fv or Fab-fragments (Hoogenboom *et al.*, Immunotechnology **4**, pp. 1-20 [1998]), other protein domains (Nygren, P.-Å. and Uhlén, M., Curr. Opin. Struct. Biol. **7**, pp. 463-469 [1997]) or linear or cyclic peptides (Cortese, R. *et al.*, Curr. Opin. Biotechnol. **6**, pp. 73-80 [1995]), for example created using combinatorial protein engineering technology; affinity pairs may also be found from such collections or libraries. Identification of variants having the desired binding capabilities may be performed using standard *in vitro* selection techniques such as phage display technology (Clackson, T. and Wells, J., Trends Biotechnol. **12**, pp. 173-184 [1994]; Smith, G.P. and Petrenko, V.A., Chem. Rev. **97**, pp. 391-410 [1997]), ribosomal display (Hanes, J. and Plückthun, A., Proc. Natl. Acad. Sci. U.S.A. **94**, pp. 4937-4942 [1997]), peptides on plasmids (Schatz, P.J., Bio/technol. **11**, pp. 1138-1143 [1993]) or mRNA-protein coupling using pyromycin (Roberts, R.W. and Szostak, J.W., Proc. Natl.

Acad. Sci. U.S.A. **94**, pp. 12297-12302 [1997]).

Proteins which form all or part of an affinity module or domain may have been made by randomisation (random mutagenesis) of a particular protein to generate ligands with novel, modified or enhanced binding characteristics. These proteins are referred to as combinatorial proteins. Such a technique typically involves random mutagenesis of a target protein, expression of the full library of these variants, e.g. on the surface of filamentous bacteriophage, followed by selection of a protein exhibiting the desired binding characteristics, this selection typically involving a binding reaction between the variant protein and an immobilised ligand (binding partner) i.e. target molecule for the protein, e.g. target analyte. The mutagenesis is random in that the resulting amino acid encoded by any particular codon is not generally pre-determined but the positions where mutations are to be introduced are generally identified in advance. The mutagenesis may involve amino acid substitution, deletion, or addition (e.g. insertion). Mutagenesis of the Z domain derived from domain B of SPA is shown in Fig. 2.

The use of an expression system such as surface display on phage provides a crucial link between genotype and phenotype; there is a self-contained unit which can be selected on the basis of its specific binding interactions and which also carries the nucleic acid encoding for the protein responsible for the observed binding characteristics. This enables identification and subsequent expression in useful amounts of the protein selected for its binding characteristics, such expression typically taking place in a transformed bacterial host.

The term 'surface display' refers to the technique involved in selection of a protein from a library of molecules which are presented (displayed) in a manner

which enables differentiation between the protein molecules on the basis of their binding characteristics. Surface display is typically performed on the surface of filamentous bacteriophage (phage display) but display can also be on the surface of bacteria, yeast cells ribosomes or using viral systems. Any "surface display" technique known or proposed in the art, may be used according to the present invention in the selection of affinity domains.

Techniques for construction of a combinatorial library of protein molecules and subsequent selection to obtain proteinaceous ligands having desired binding characteristics are known in the art (Nygren, P. and Uhlén, supra (1997) 7: 463-469).

Generally, a protein molecule, perhaps having intrinsic beneficial properties such as temperature or pH insensitivity, is used as a scaffold and a combinatorial library is then constructed via random but targeted amino acid substitutions (or other mutations) of that protein molecule, in order to produce a library of molecules having different binding characteristics. Surface residues are generally targeted for random mutagenesis.

Suitable protein scaffolds may simply be linear peptides but preferably the scaffold will possess a folded three dimensional structure which has the potential for higher affinities and is less susceptible to proteolytic degradation. Rather than designing a scaffold *de novo*, naturally existing proteins or domains are usually selected for further engineering. The choice of protein scaffold depends on several parameters including an ability to be effectively expressed in a desired host organism e.g *E. coli* when the randomised protein is to be displayed as a fusion protein with a filamentous phage coat protein. The protein should also comprise sufficiently large regions on its surface which are tolerant to substitution (or insertion or deletion

etc.) without losing the overall three dimensional structure. If the library is to be produced synthetically, a small overall size is a prerequisite. Where the selected scaffold protein has a binding function, amino acid residues involved in that interaction may be a target for randomisation. Randomisation may be performed in order to enhance known binding properties or to develop ligands with new specificities.

Suitable scaffold molecules are discussed in Nygren et al. (1997, supra) and include cyclic peptides having 40 or more residues in a constrained sequence, immunoglobulin-like scaffolds including Fv or single-chain (scFv) domains, bacterial receptors such as the 58-residue one-domain *Staphylococcal* protein A (SPA) analogue Z (the "Z Domain" being a derivative of the B domain of SPA), or other domains or analogues of SPA, DNA-binding proteins particularly zinc fingers and protease inhibitors. All these molecules can be stabilised by substituting one or more native asparagine residues with a less alkaline sensitive amino acid before a combinatorial library of the pre-stabilised protein is made.

Of particular interest is the bacterial receptor domain Z (Nord, K., Nilsson, J., Nilsson, B., Uhlén, M. and Nygren, P. Protein Engineering (1995) 8, 6, 601-608). This paper by Nord *et al.* describes a suitable method of constructing a combinatorial library of protein molecules which can be applied to a range of scaffold molecules. The method described is solid-phase-assisted and based on the stepwise assembly of randomised single-stranded oligonucleotides.

Selection from a generated protein library can be performed in a number of different ways known in the art, including bead immobilised libraries (McBride, J.D., Freeman, N., Domingo, G.J. and Leatherbarrow, R.J., J. Mol. Biol. [1996] 259: 819-827), fusions to

DNA-binding proteins (Schatz, P.J., Biotechnology [1993] 11, 1138-1143) and when displayed on bacteria (Lu, Z., Murray, K.S., Van Cleave, V., LaVallie, E.R., Ståhl, M.L. and McCoy, J.M. Biotechnology [1995], 13 366-372) or phage (Clackson, T. and Wells, J. TIBTECH (1994) 12, 173-183) as well as yeast cells, (Boder, E.T. and Wittrup, K.D., Nature Biotechnology (1997) 15, 553-557) and in viral systems (Ernst, W., Grabher, R., Wegner, D., Borth, N., Graussauer, A. and Katinger, H. Nucleic Acids Research (1998) 26, 1718-1723 and Grabher, R., Ernst, W., Doblhoff-Dier, O., Sara, M. and Katinger, H. BioTechniques (1997) 22, 730-735).

International Patent Application, publication No. WO 95/19374 describes the generation of a combinatorial library of Z-variants, see in particular Example 4.

Thus in a further aspect, the present invention provides a method of preparing a combinatorial library of protein molecules, molecules selected from said library being for use in an affinity module as defined above. The invention also provides the use of a molecule selected from a molecular library as an affinity domain in a self-assembled biomolecular structure as described herein.

In a yet further aspect of the present invention is provided a method of selecting a molecule for use in an affinity module (i.e. as an affinity domain) as defined above, wherein said molecule is selected from a molecular library on the basis of its ability to bind to a particular target molecule. The 'target molecule' will typically comprise an affinity domain used in the self-assembly of a biomolecular structure, the selected molecule forming an affinity pair therewith.

Albumin Binding Protein (ABD), a protein domain with affinity towards human serum albumin (HSA) may be used as an affinity domain or as a further, linker component of the affinity module which is attached to affinity domains. It is derived from a cell wall

anchored bacterial receptor protein from *Streptococcus* G148. The wild type sequence of this protein has four asparagine residues, enhanced stability is observed when just one of these residues is substituted but preferably all four residues are replaced by less alkaline sensitive residues. The wild type amino acid sequence of ABD is:

LAEAKVLA**N**RELDKYGV-SDYYK**N**LI**NN**AKTVEGVKALIDEILAALP

the asparagine residues have been indicated in bold. The hyphen simply indicates that other molecules in the same family have an additional amino acid in this position.

ABD may be subjected to randomisation to create a protein (i.e. a combinatorial protein) having e.g. modified binding characteristics for HSA and/or that is able to bind any target of choice and also retains affinity towards HSA.

The affinity modules may be synthesised by recombinant DNA technology, by transformation of a suitable host cell with a plasmid or viral construct encoding for the affinity module protein, or by chemical synthesis.

Chemical or genetic fusion may be employed as necessary to link affinity domains so as to form appropriate bi- or multi-valent or bi- or multi-functional affinity modules. Some modules such as the 5 domain SPA will be expressed as a whole 'module' without the need to introduce regions into the nucleotide sequence which will provide linking regions or to modify the expressed protein to join the domains together. Alternatively, Example 3 describes how a genetic construct may be prepared to encode a fusion protein which can act as an affinity module.

Typically, the affinity between modules (i.e. between the domains thereof) should be in the range of $K_{aff}$ = $10^4$-$10^{11}$ $M^{-1}$ in order to allow stable self-assembled

structures to form under a range of conditions, for example under physiological conditions or at different pHs and/or temperatures. For certain applications it may, however, be advantageous to prepare structures which may be reversibly disassembled as a result of, for example, pH or temperature change.

The nature of self-assembled biomolecular structures formed in accordance with the invention may, for example, be investigated by techniques such as size exclusion chromatography, electron microscopy, viscosity analyses, microcalorimetry, nuclear magnetic resonance spectroscopy, atomic force microscopy and X-ray crystallography.

In a further aspect, the present invention provides a proteinaceous affinity module which comprises two or more proteinaceous affinity domains which may be the same or different, each affinity domain having specific and exclusive affinity for a given binding partner, wherein at least one of the affinity domains has been selected from a molecular library. The use of such modules in the formation of a self assembled biomolecular structure constitutes a further aspect of the present invention. Nucleic acid molecules encoding such a protein as well as cells expressing the protein constitute further aspects of the present invention.

As discussed above, 'molecular library' refers to a combinatorial protein library generated by mutagenesis of a starting molecule or the components of 'library' of molecules harvested from natural sources, a naive library as described by Hoogenboom et al. supra.

Preferably, one or more of the affinity domains is capable of specific and exclusive interaction with a bacterial receptor domain such as SPA. Preferably one of the affinity domains is derived from a bacterial receptor domain such as domain Z. The inventors have isolated particularly useful affinity domains which are derived from domain Z and are identified herein as $Z_{8G}$

and $Z_{6s}$; these protein molecules constitute further aspects of the present invention. The amino acid sequences of these affinity domains are given in Fig. 2. Also included within the scope of the present invention are variants of these molecules which differ only in the regions of the molecule which have not been randomized (i.e. not marked *) and wherein those modifications do not significantly affect the binding affinity of the domain as a whole for SPA. Modifications may include additions, substitution or deletion. Preferably, any modified domains will have at least 80%, more preferably at least 90% of the binding affinity of the $Z_{8G}$ or $Z_{6s}$ molecule for SPA. Preferred affinity modules comprise affinity domains linked to ABD.

Nucleic acid molecules encoding the affinity domains and affinity modules described above, including vectors incorporating such nucleic acid and cells expressing such domains and modules constitute further aspects of the present invention. Methods of selecting affinity domains, for example by phage or ribosome display, for, a molecular library constitute a still further aspect of the present invention.

Also, a method of producing a self-assembled biomolecular structure as defined herein, which comprises admixing affinity modules in an environment which enables interaction between the affinity domains thereof comprises a further aspect of the present invention.

The structures have a wide range of therapeutic diagnostic and other applications. These structures which may be in the form of fibres, sheets or three dimensional structures may be useful as materials, e.g. for encapsulation of active agents or they may have a more active functional role, e.g. in bioelectronic applications. A variety of applications are discussed below.

Chemical groups and/or molecules (including further

proteins) may if desired be incorporated into the affinity modules, for example by chemical coupling or genetic fusion, thereby permitting the preparation of assembled structures with a variety of properties and functionalities. Structures obtainable in accordance with the invention may thus have a wide range of applications, including, for example diagnostic applications, e.g. to obtain highly avid reagents; clinical applications, e.g. to obtain controlled delivery of therapeutics; nano-fabrication applications, e.g. to obtain spontaneous build up of ordered small-scale structures; biotechnological applications, e.g. to obtain thermally or chemically reversible protein networks; and provision of basis for stepwise enzymatic treatment of a substrate at a defined position.

Thus, for example, in the formation of highly avid diagnostic reagents, effector molecules such as reporter enzymes or other detection molecules, including alkaline phosphatase, β-galactosidase, horse radish peroxidase, fluourescein isothiocyanate, green fluorescent protein or derivatives thereof, luciferase and/or additional affinity reagents such as biotin, streptavidin, cellulose binding domains, zinc fingers, antibody fragments or protein or peptide moieties may be chemically or genetically coupled to a suitable affinity module.

In clinical applications enzymes or therapeutics (including vaccines) may be encapsulated in self-assembling biomolecular structures by mixing the affinity modules with the substance to be encapsulated. Controlled release of the active substance may occur upon, for example, a change in buffer or temperature conditions, or *in vivo* upon spontaneous disassembling of the structures. An advantage of the *in vivo* use of self-assembling biomolecular structures is that these are normally biodegradable.

Self-assembling biomolecular systems in accordance

with the invention may be used to provide controlled "plugging" of cavities or holes, either *ex vivo* or *in vivo*, for example by controlled release of enzymes or other active substances. Thus, for example, protein self-assembly systems capable of preventing bleeding may be prepared by derivatisation of affinity modules with substances directing the assembly to specific sites around injuries.

Self-assembled protein structures according to the invention may be used in the preparation of hydrogels capable of reversible gelation under selected buffer or temperature conditions, and in the preparation of protein fibres or films which may be used to form biodegradable materials such as threads or cloths.

Following address-specific deposition of a suitable anchoring affinity module onto a bio-chip, higher forms of structures may be produced in a stepwise manner using further affinity modules in accordance with the invention. Such self-assembly may be used to create structures for use in the detection of or functional studies of biological samples. In addition, through derivatisation of affinity modules with suitable substances such as enzymes or organic substances, a build up of complexes mimicking natural systems, for example substrate conversion by concerted action of several enzymes, or production of energy or light as in photosynthesis or photorespiration, may be accomplished.

In bioelectronic applications the spontaneous assembly of affinity modules into ordered structures may, for example, serve as a means to create small circuits useful for conducting electric charges or currents. Suitably derivatised affinity modules, e.g. highly charged modules or modules into which molecular groups with high electron densities have been directly or indirectly incorporated, may be used to prepare self-assembled structures useful as "wires" for nano-fabrication of electric circuit components such as

resistors, transistors or capacitors. Self-assembled structures may also be used for site-specific deposition of metallic materials such as silver or gold, for example in order to create anchoring points from which a circuit may be built up by attachment of appropriate affinity modules, for example using affinity modules containing cysteine residues to couple to deposited gold.

The invention will now be further described by the following non-limiting examples in which:

**Figure 1** shows biosensor analysis of selected staphylococcal protein A-specific affibodies. In the analysis, affibodies were analyzed as fusion proteins to the serum albumin binding domain ABD. Upper panel: (A) Response curve obtained for injection of the 8G affibody over a protein A-coated surface; (B) Response curve obtained for injection of a control affibody (non-protein A binding) over a blank surface; (C) Response curve obtained for injection of a control affibody (non-protein A binding) over a protein A-coated surface. Lower panel: (A) Response curve obtained for injection of the 6S affibody over a protein A-coated surface; (B) Response curve obtained for injection of the 6S affibody over a blank surface; (C) Response curve obtained for injection of a control affibody (non-protein A binding) over a protein A-coated surface.

**Figure 2** provides the sequences of selected anti-staphylococcal protein A affibody affinity modules 8G (SEQ ID No. 2) and 6S (SEQ ID No. 3). Amino acid sequences are shown in one-letter code. Also shown is the amino acid sequence of the wild type Z domain (SEQ ID No. 1) used as scaffold for library constructions. In addition, boxes corresponding to the three alpha-helical regions of the domain are shown. Randomized positions are shown as asteriscs.

**Figure 3** shows SDS-PAGE analysis of proteins purified by protein AG or HSA-affinity-chromatography

from periplasmic preparations of a culture of *E. coli* cells harboring the 8G clone phagemid encoding the $Z_{8G}$-ABD fusion protein. Lane 1: HSA-affinity purified material; Lane 2: HSA-affinity purified material (1:10 dilution); Lane 3: Protein AG-affinity purified material; Lane 4: Protein AG-affinity purified material (1:10 dilution). In both cases, the $Z_{8G}$-ABD protein is purified to near homogeneity, showing upon selective interactions between the $Z_{8G}$ moiety and protein A and the ABD moiety and HSA, respectively. Molecular weight markers are: 14, 20, 30, 43, 66, 92 kDa.

**Figure 4** shows construction and production of the divalent staphylococcal protein A-binding hybrid affinity module ($Z_{8G}$-ABD-$Z_{8G}$). The figure shows the cloning chart for the joining of two different PCR products by splice overlap extension (SOE) PCR to result in the gene encoding the $Z_{8G}$-ABD-$Z_{8G}$ protein.

**Figure 5** shows SDS-PAGE analysis of purified $Z_{8G}$-ABD-$Z_{8G}$ protein. The figure shows the results from an SDS-PAGE analysis of HSA-affinity chromatography purified protein (lane 1) or protein AG-purified protein (lane 2) from a periplasmic extract of *E. coli* cells harboring the pE-$Z_{8G}$-ABD-$Z_{8G}$ construct. Molecular weight markers are: 14, 20, 30, 43, 66, 92 kDa.

**Figure 6** shows native PAGE analysis of complex formation between $Z_{8G}$-ABD-$Z_{8G}$ protein and staphylococcal protein A. **Lane 1:** $Z_{8G}$-ABD-$Z_{8G}$ protein; **Lane 2:** protein A protein; **Lane 3:** An aliquote from a mixed sample containing 10 $\mu$M of $Z_{8G}$-ABD-$Z_{8G}$ and 2.5 $\mu$M of protein A; **Lane 4:** An aliquote from a mixed sample containing 10 $\mu$M of $Z_{8G}$-ABD-$Z_{8G}$ and 5 $\mu$M of protein A; **Lane 5:** An aliquote from a mixed sample containing 10 $\mu$M of $Z_{8G}$-ABD-$Z_{8G}$ and 7.5 $\mu$M of protein A; **Lane 6:** An aliquote from a mixed sample containing 10 $\mu$M of $Z_{8G}$-ABD-$Z_{8G}$ and 10 $\mu$M of protein A.

**Figure 7** provides atomic force microscopy images from analyses of complex formation between $Z_{8G}$-ABD-$Z_{8G}$

protein and staphylococcal protein A. Left panel: Analysis of a dried 20 nM sample of protein A-cys by AFM. Right panel: Analysis of a dried 10 nM sample of protein A by AFM, in which the divalent protein A-binding affinity module $Z_{8G}$-ABD-$Z_{8G}$ had previously been added to a final concentration of 20 nM.

## EXAMPLES

**Standard procedures:**

Phagework including preparation of phage stocks and titration was performed as described above (Nord et al. (1997)). The phage libraries Zlib-1 and Zlib-2 have also been described above (Nord et al., 1997). Standard cloning work was performed as described in (Sambrook, J., Fritsch, E. F. and Maniatis, T. Molecular cloning: a laboratory manual, 2nd edn., Cold Spring Harbor Laboratory, New York, 1989). Native gel protein electrophoresis was performed in a Novex Xcell II (San Diego, CA) as described by the supplier. PCR amplifications were performed using standard conditions using oligonucleotide primers from Table 1 as specified in the examples.

**Table 1. List of oligonucleotide primers.**

| Name | Sequence 5´-3´ |
|---|---|
| SNAP-1 | GCG AAT TCG GTA GAC AAC AAA TTC AAC AAA G |
| SNAP-2 | AGG TAA TGC AGC TAA AAT TTC |
| SNAP-3 | GAT GAA ATT TTA GCT GCA TTA CCT AGC GGC AGC GTA GAC AAC AAA TTC AAC AAA G |
| SNAP-4 | GCG CGC TGC AGT TAG TCG ACT TTC GGC GCC TGA GCA TCA T |
| RIT27 | GCT TCC GGC TCG TAT GTT GTG TG |
| RIT28 | AAA GGG GGA TGT GCT GCA AGG CG |
| USP | CGT TGT AAA ACG ACG GCC AG |

| | |
|---|---|
| RSP | TTC ACA CAG GAA ACA GCT ATG ACC |
| NOKA-2 | CGG AAC CAG AGC CAC CAC CGG |

---

*E. coli* cells used for phage work, cloning and expression were either RR1ΔM15 (Rüther, U. Nucl. Acids Res. 10: 5765-5772, 1982). Osmotic shock procedures were performed as described by (Nygren et al.,1988. Affinity chromatography purifications of proteins on IgG or HSA-Sepharose resins were also performed as described above (Nygren et al., J. Mol. Recognit. 1:69-74, 1988). Human polyclonal IgG was supplied by Pharmacia and Upjohn AB, Stockholm.

DNA sequencing was performed by cycle sequencing (Carothers et al., BioTechniques 7:494-499, 1989) using ThermoSequenase DNA polymerase (Amersham Pharmacia Biotech) and primers as indicated. Sequencing reactions were loaded onto a MegaBace 1000 instrument (Amersham Pharmacia Biotech, Uppsala, Sweden) and the results were analyzed using Sequencher 3.1.1 software.

Atomic force microscopy (AFM) was performed using a Nanoscope IIIa (Multimode) from Digital Instruments. Protein containing solutions were dried onto newly prepared mica surfaces and analyzed without further treatment.

Biosensor experiments were performed using a Biacore 2000 instrument (Biacore AB, Uppsaal, Sweden) according to the suppliers recommendations. Sensorgrams were analyzed using Biaevaluation 2.1 software. CM-5 sensor chips were used for coupling of proteins via amine cooling chemistry according to the suppliers recommendations.

**Example 1. Phage display selection of a staphylococcal protein A-specific binding protein.**

In order to isolate binding proteins capable of selective binding to staphylococcal protein A, phage display in vitro selection experiments were performed

using protein libraries constructed by combinatorial engineering of a protein domain (Z), derived from staphylococcal protein A (Nord et al. 1997). The libraries Zlib-1 and Zlib-2 have been described previously (Nord et al., 1997) and were constructed by randomization of thirteen surface located positions of the three-helix bundle domain Z. For the selection of binding proteins to the protein A, phage stocks from the libraries were prepared as described by Nord and coworkers (Nord et al., 1997). As targets during selections, recombinant protein A was used (Amersham Pharmacia Biotech, Uppsala, Sweden).

Approximately 570 μg of the purified protein A protein was subjected to in vitro biotinylation using a commercial biotinylation kit (Pierce, art. No 21335) based on amine coupling chemistry. For selections, approximately 40 μg of the biotinylated protein A was immobilized onto streptavidin coated paramagnetic beads (Dynal M280-SA, Dynal AS, Norway) according to the suppliers recommendations. This resulted in an immobilization of approximately 5 μg protein A per mg of beads. In each round 5 mg of beads were used.

Selections were performed for five cycles using conditions for phage stock preparation as described earlier (Nord et al. 1997), with the addition of a preincubation step of phagestocks with streptavidin-coated beads (containing no biotinylated proteins) as to prevent the selection of streptavidin binding Z variants. In each round of selection, approximately $10^{11}$ cfu of the phagestocks (a 50/50 mixture of Zlib-1 and Zlib 2-phage libraries) were used as input to approximately 5 mg streptavidin coated beads in final a volume of approximately 100 μl. This resulted in a final target concentration of approximately 7 μM protein A.

In the selection rounds, either low pH (0.1 M Glycin-HCl, pH 2.2) (S-indexed binders) or competition

with a 100-fold molar excess of human polyclonal IgG (300 μM) (G-indexed binders) was used for elution. After five cycles of selections, aliquots of the respective eluates were used to infect RR1ΔM15 cells which were spread on selective media (ampicillin agar plates, 100 μg/ml).

**Example 2. Characterization of a protein A-binding affibody protein.**

To investigate if resulting clones from the selection experiment corresponded to affibody variants capable of selective protein A binding, eight randomly picked clones were studied further. Cells harboring the different phagemid construct were cultivated in 100 ml shakeflasks in a tryptone soy broth growth medium (TSB, Difco) containing 100 μg/ml of ampicillin for production of selected affibody proteins as fusions to a serum albumin binding domain from streptococcal protein G, as encoded from their respective phagemid vectors. Proteins were purified by HSA-affinity chromatography (Nygren et al., 1988) from samples prepared by osmotic shock of cells to release periplasmic material.

Purified affibody-ABD fusion proteins were subsequently analyzed for protein A-binding using biospecific interaction analysis (BIA) in a biosensor instrument (Biacore 2000, Biacore AB, Uppsala, Sweden). To a flow cell surface of a CM-5 sensor chip approximately 1200 RU protein A was coupled using amine coupling chemistry (NHS/EDC) according to the suppliers recommendations. Injections of samples of the different affibody-ABD fusion proteins resulted in the identification of two clones showing specific protein A-specific binding (Figure 1). These two clones, 8G and 6S, had been selected using competitive human IgG (8G) and low pH-elution (6S), respectively. The affibody genes in the phagemids of the two positive clones 8G and 6S were subjected to DNA sequencing by cycle DNA

sequencing using primers RIT27 and NOKA-2 for PCR amplification and primer NOKA-2 for sequencing reactions with dye-labeled dideoxy nucleotide terminators. The results from this sequencing is shown in figure 2, in which deduced amino acid sequences of the two affibody clones are shown. This result showed that affinity proteins showing specific binding to the highly repetitive staphylococcal protein A could be selected from the Z domain libraries.

The selectivity in binding of the 8G clone was further demonstrated in the purification of the $Z_{8G}$-ABD fusion protein by protein AG-affinity chromatography (Jansson et al., FEMS Immunol. Med. Microbiol. 20:69-78, 1998) from an osmotic shock preparation of a culture of RR1ΔM15 cells harboring the corresponding phagemid vector. Analysis of low pH-eluted proteins by SDS-PAGE (Figure 3, lane 1) showed upon a single band corresponding to the $Z_{8G}$-ABD fusion protein (molecular weight=13.5 kDa). The same protein was also purified using HSA-affinity chromatography, employing the ABD moiety of the fusion protein (Figure 3, lane 2). In both cases, a product of high purity was obtained showing that both moieties in the fusion protein are functional and bind their cognate binding partner with high selectivity when genetically linked as a fusion construct.

**Example 3. Construction of a divalent protein A-specific hybrid affinity module.**

A genetic construct was assembled encoding a fusion protein $Z_{8G}$-ABD-$Z_{8G}$, containing a serum albumin binding domain (ABD), flanked both N-and C-terminally by the 8G affibody domain. Thus, this protein construct corresponds to a divalent protein A-binding protein. The construct was assembled using splice overlap extension (SOE) PCR from two different PCR products (Figure 4) sharing common sequence corresponding to the

intended junction. The first PCR fragment (Fragment A, figure 4) was produced by PCR amplification using primers SNAP-1 and SNAP-2 on a template consisting of phagemid vector pKN1-$Z_{8G}$ resulting in a PCR product encoding the $Z_{8G}$ and ABD domains. Similarly, the second PCR product (fragment B, figure 4) was also prepared by PCR using the pKN1-$Z_{8G}$ phagemid as template but with the use of primers SNAP-3 and SNAP-4 resulting in a PCR product encoding the $Z_{8G}$ affibody domain. The two fragments were subsequently joined through overlapping sequences in a SOE-PCR reaction, using approximately 5 ng each of the two PCR products. For the amplification of the resulting joined fragment, 5 mole of outer PCR primers SNAP-1 and SNAP-4 were added to the PCR reaction tube before temperature cycling. The polymerase used for this procedure was *Taq* DNA polymerase (Perkin Elmer Corp.). The temperature cycling scheme for this procedure was: 96°C (30 sec), 40°C (30 sec) and 72°C (45 sec) for ten cycles followed by 25 cycles of 96°C (15 sec), 55°C (30 sec) and 72°C (45 sec).

The resulting PCR product was ligated into a pGEM cloning vector (AT-cloning, Clontech) and the ligation mixture transformed to RR1ΔM15 cells. Resulting transformants (ampicillin containing agar plates) were PCR-screened using primers RIT 27 and RIT 28 (Hultman et al., Nucl. Acids Res. 17:4937-4946, 1989). and sequenced using primers USP and RSP (Amersham Pharmacia Biotech). A clone with the correct sequence of the insert was identified and purified plasmid DNA from this clone was cleaved with restriction enzymes *Pst* I and *Eco* RI. The purified gene fragment encoding the divalent affinity module $Z_{8G}$-ABD-$Z_{8G}$ was subsequently ligated into the expression vector pE318 (Jendeberg et al., J. Immunol. Methods, 201:25-34, 1997) cleaved with the same enzymes and the ligation mixture transformed to *E. coli* cells. One clone (no. 18) with a correct size of the insert as judged from agarose gel electrophoresis after cleavage

of purified plasmid DNA was subjected to confirmatory DNA sequencing and was found to be correct. This clone was subsequently used for production of protein $Z_{8G}$-ABD-$Z_{8G}$ protein which was purified by either or protein AG or HSA-affinity chromatography (Nygren et al., 1988) from a periplasmic extract of an overnight culture of *E. coli* cells harboring the pE-$Z_{8G}$-ABD-$Z_{8G}$ construct. The successful purification of the $Z_{8G}$-ABD-$Z_{8G}$ fusion protein by the two principles HSA or protein AG-affinity chromatography showed that the ABD domain flanked at both the N-terminal and C-terminal by $Z_{8G}$ domains as well as the $Z_{8G}$ domains were functional in the fusion protein. Analysis by SDS-PAGE showed that the majority of the protein were of full length with a strong band of the apparent molecular weight of 19.7 kDa (Figure 5).

**Example 4. Analysis by native gel electrophoresis of self-assembled protein complexes.**

To investigate if the two affinity modules $Z_{8G}$-ABD-$Z_{8G}$ and protein A were capable of forming protein complexes of higher order, the two proteins were mixed at different molar ratios and subsequently analyzed by native polyacrylamide gel electrophoresis (PAGE). In concordance with a productive complex formation between the two affinity modules, a band corresponding to a larger form showing slower migration appeared in samples in which the two proteins had been mixed at different stoichiometrical ratios (Figure 6, lanes 3-5). The appearance of this band indicated that a bimolecular complex between the $Z_{8G}$-ABD-$Z_{8G}$ and the protein A affinity modules had been formed, since the intensity of the smaller molecular weight band of $Z_{8G}$-ABD-$Z_{8G}$ decreased with the increasing intensity of the larger band. At even higher relative concentrations of protein A, bands corresponding to even slower migrations appeared, suggesting that at these ratios, higher order networks (more than two affinity modules) had been spontaneously

formed from biospecific interactions which were stable enough to hold together even when analyzed by the gel electrophoresis procedure.

**Example 5. Analysis by atomic force microscopy of self-assembled protein complexes.**

Analyses by atomic force microscopy (AFM) were performed to investigate if complexes formed between the two affinity modules $Z_{8G}$-ABD-$Z_{8G}$ and protein A could be visualized. Mica surfaces were prepared and cleaned with alcohol and used to apply samples of proteins. Two different samples were analyzed in a Nanoscope IIIa AFM instrument (Digital Instruments) which was used to scan the surfaces in air. Instrumental settings were: Z-height=2 nm; scan rate=1.5 Hz; tapping mode). In a first experiment, 20 $\mu$l of a pure 20 nM Protein A solution was applied onto the mica surface and was allowed to dry in air. Analysis of this sample by AFM revealed spatially distributed spots with an apparent mean diameter of approximately 20 nm, indicating that protein A molecules were evenly distributed on the mica surface (Figure 7, left panel). In contrast, when a second sample consisting of a mixture between protein A protein (10 nM) and the divalent protein A-binding $Z_{8G}$-ABD-$Z_{8G}$ affinity module (20 nM) was analyzed with the same instrumental settings, larger protein forms were clearly visible (Figure 7, right panel). This indicates that the addition of the $Z_{8G}$-ABD-$Z_{8G}$ affinity module capable of selective biospecific interaction with the five-domain protein A affinity module resulted in the spontaneous formation of multi protein complexes of higher order structures from building blocks held together by biospecific interactions.